

高中生教育議題之討論-以Dcard討論板為例

第二組成員：許哲維、鄒志宇、曾裕堯、周立揚

報告日期：2022/7/30

目錄

1

教育議題

2

數據來源

3

分析方法

4

分析流程

5

結果與討論

6

結論與建議

7

參考資料

教育議題

- 高中生教育制度的改變



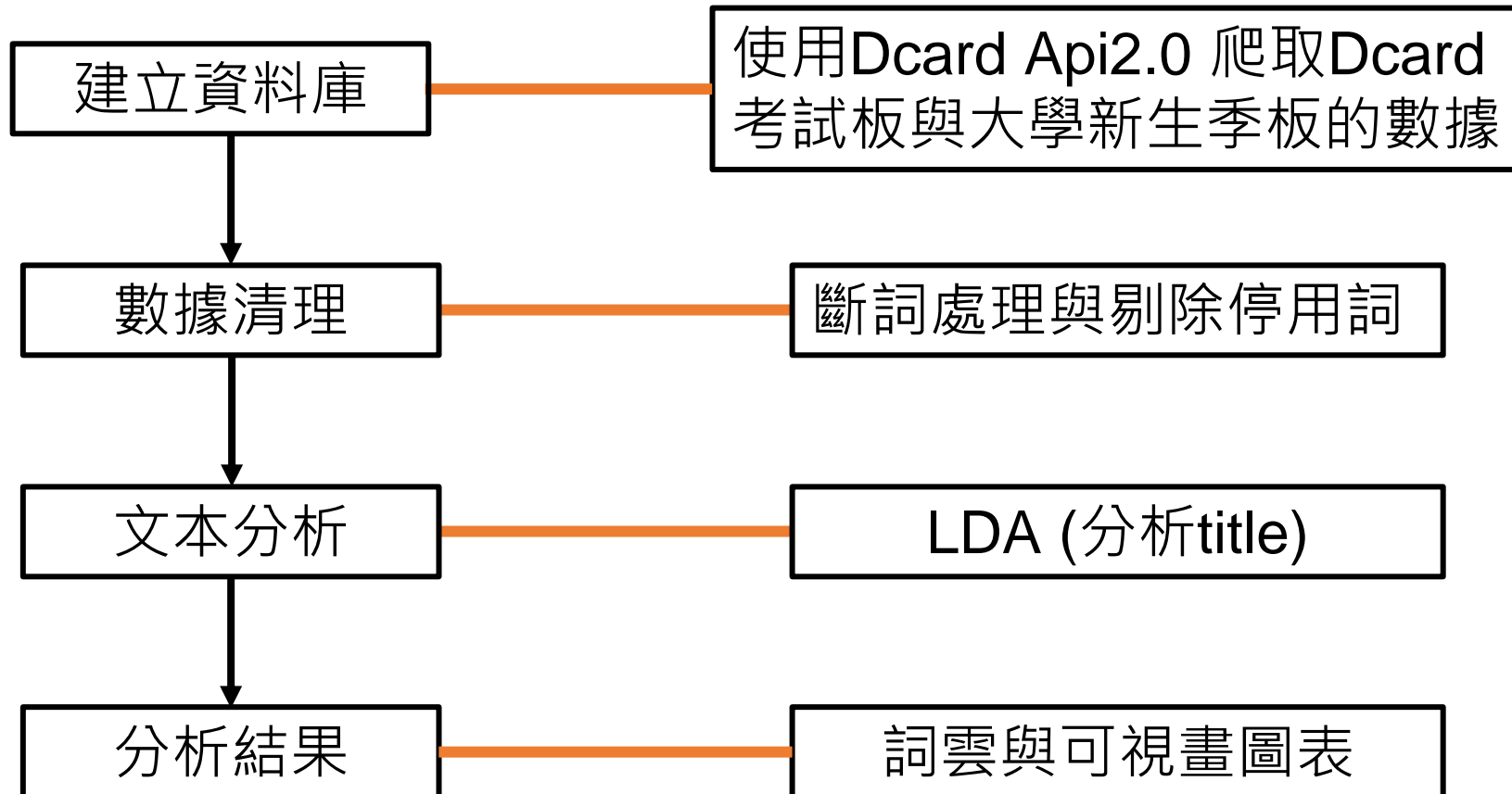
數據來源

社群媒體	Dcard	
板塊	考試板	大學新生季板
數據量	9909	46300
年份	2021/12/6-2022/7/25	2018/7/28-2022/7/28

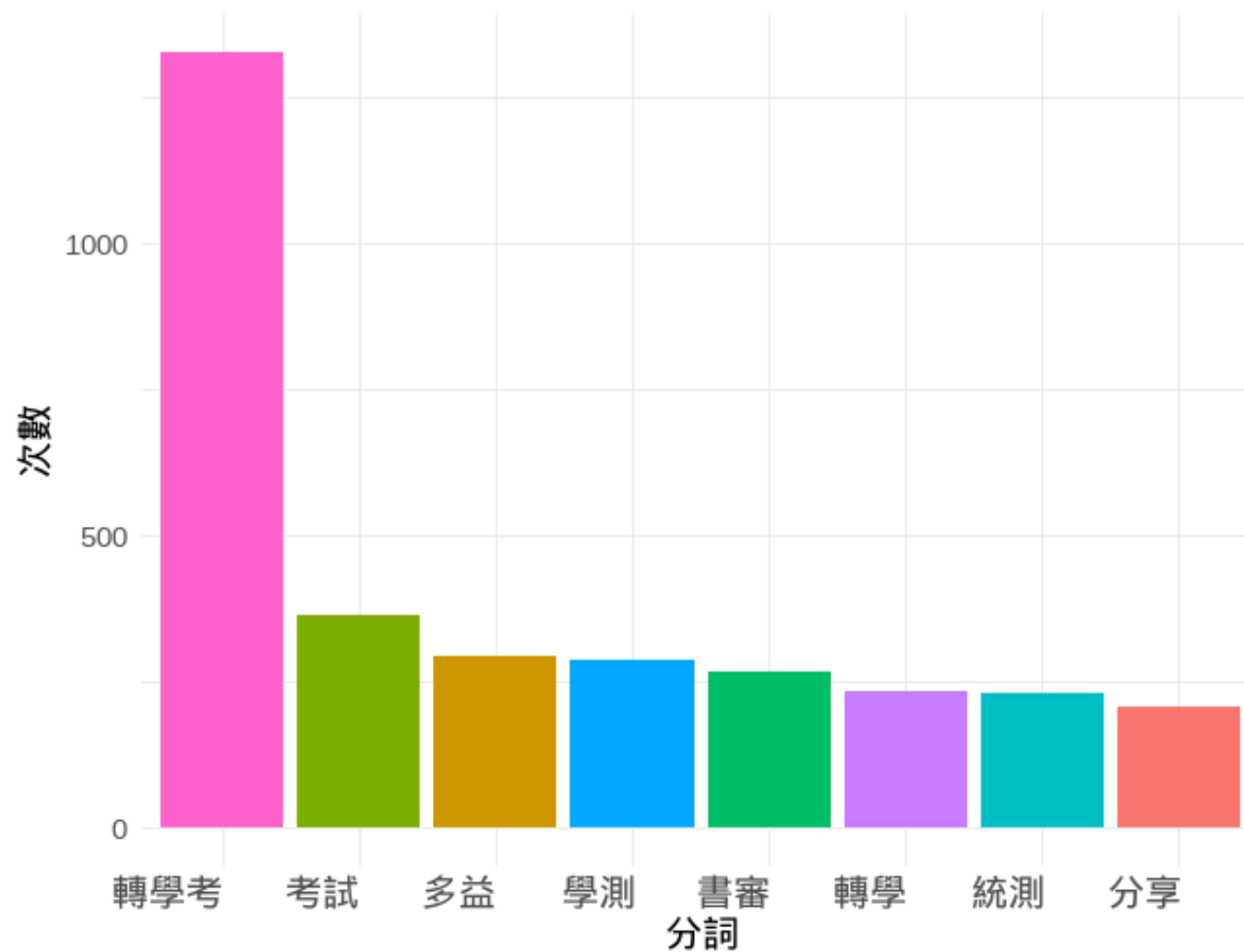
分析方法

		R語言	
斷詞處理	字典、停用詞	NLP	Topic Model (LDA)
jieba套件	自行建立	Quanteda套件	STM套件

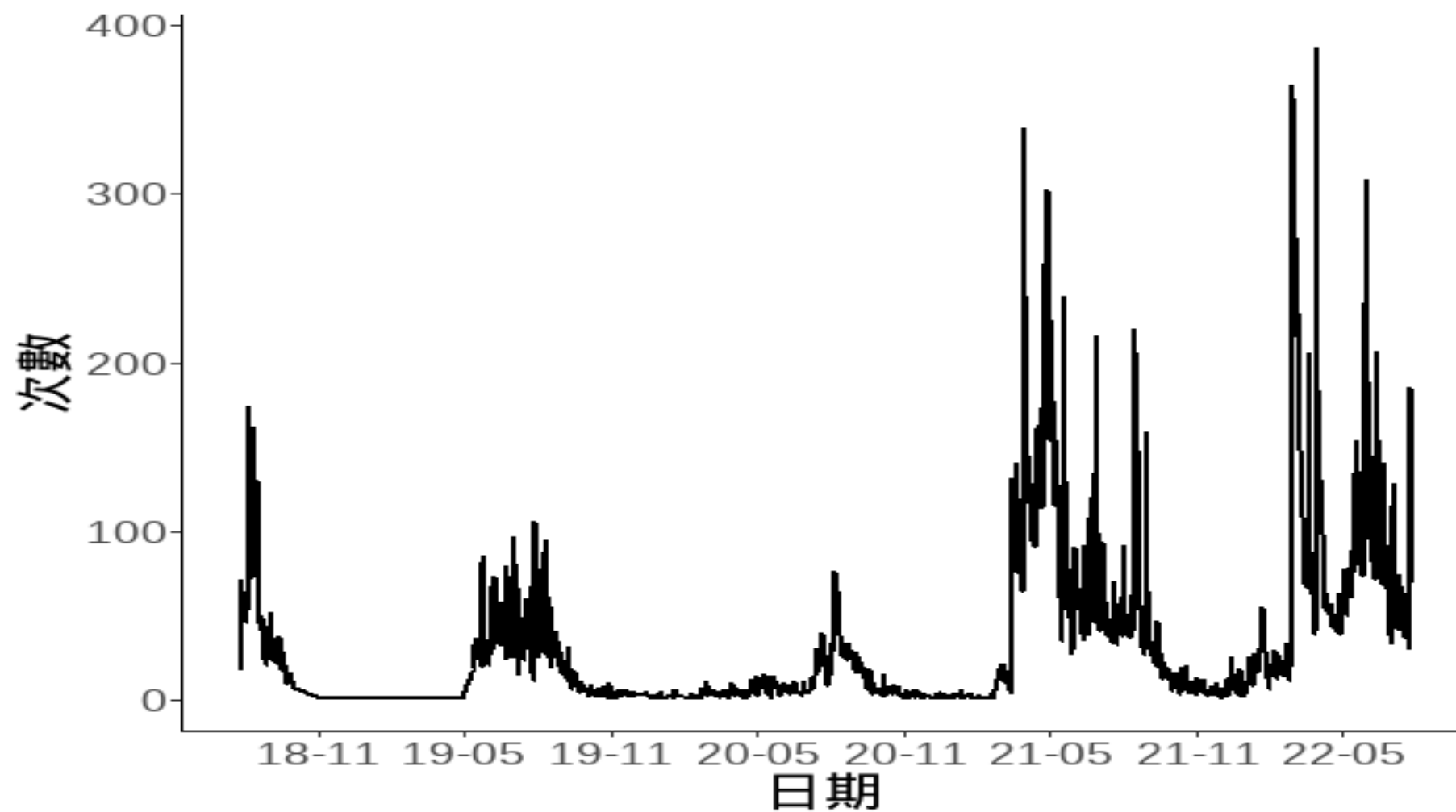
分析流程



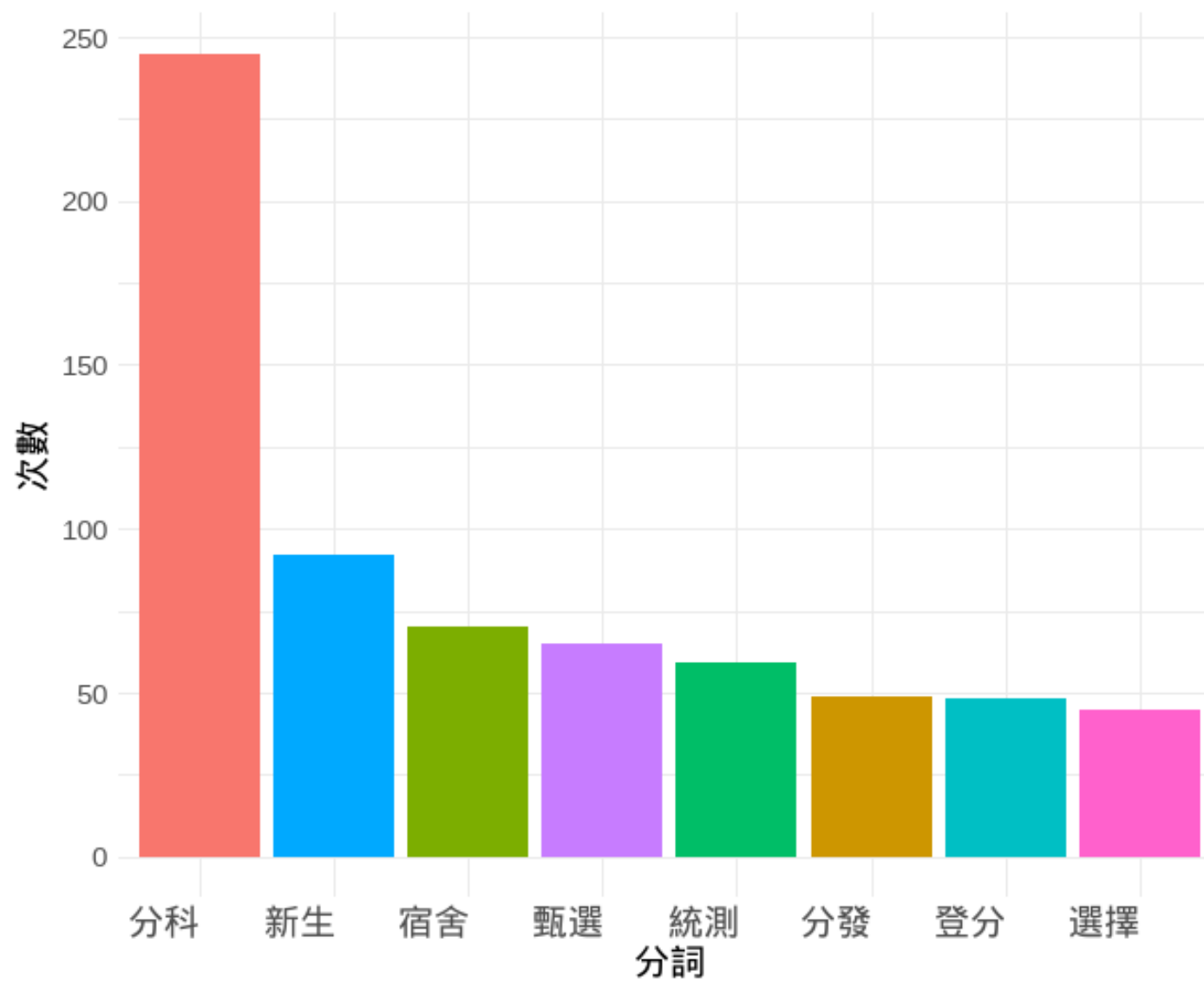
考試板之標題最常見字詞



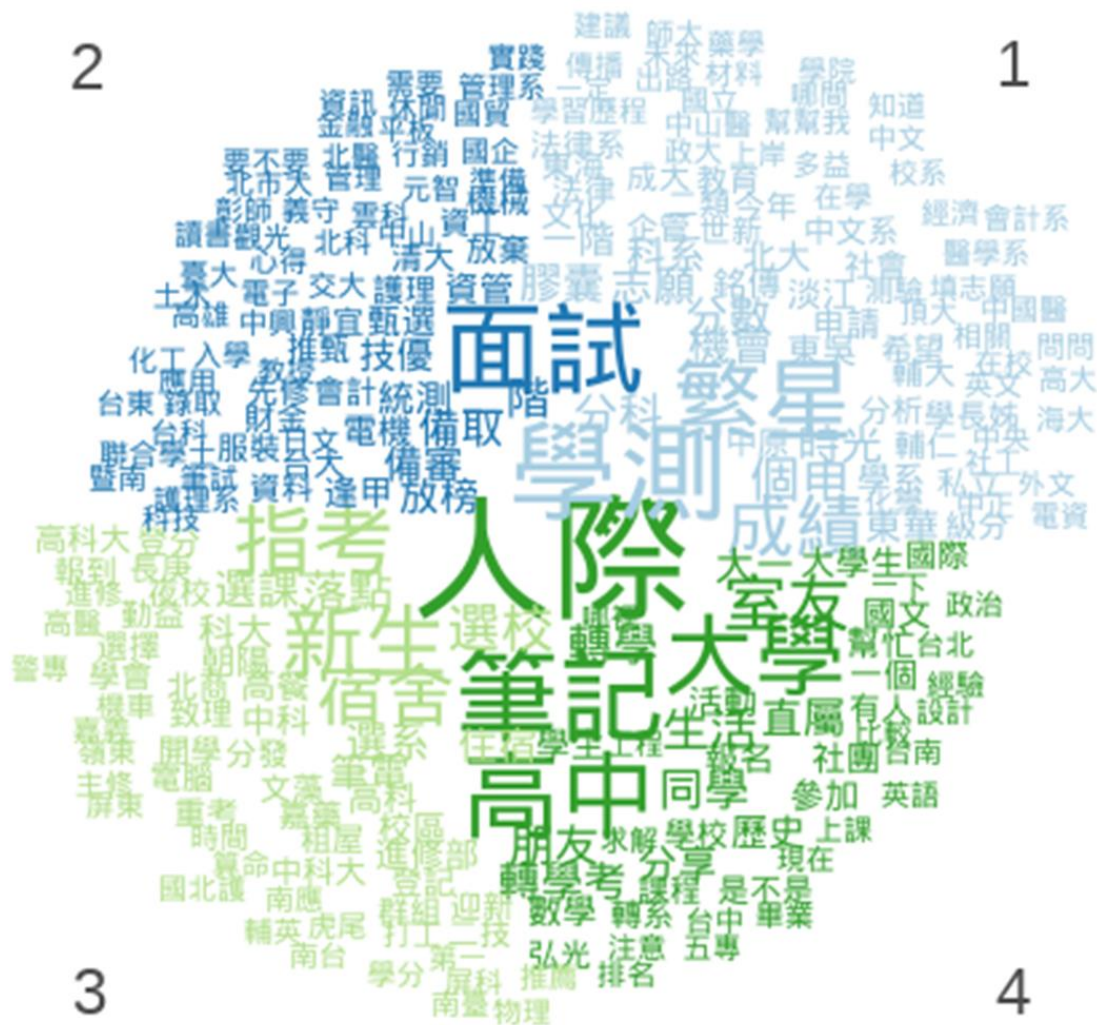
大學新生季板之發文次數的變化



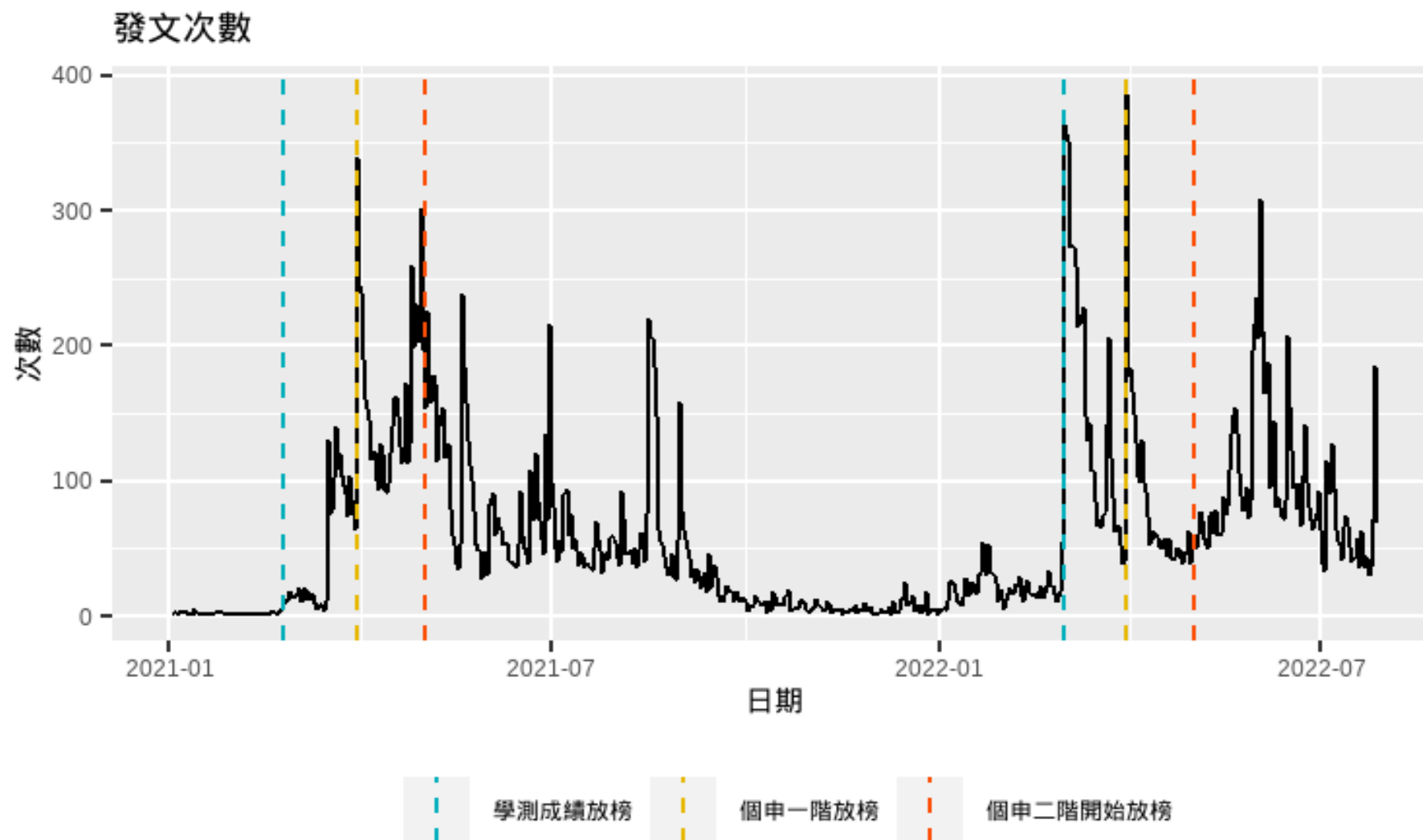
標題之最常見字詞



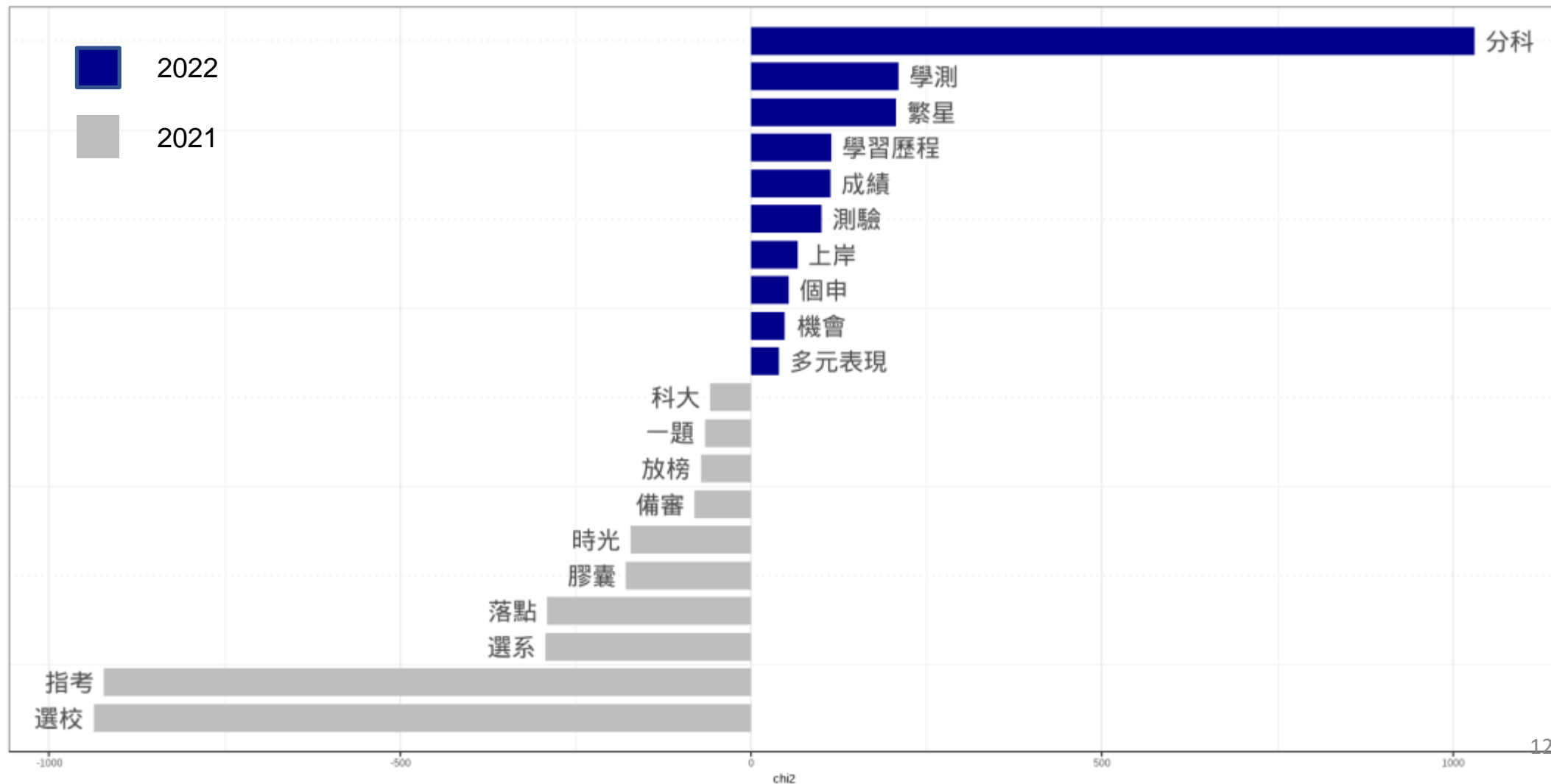
詞雲之各季度熱門關鍵詞



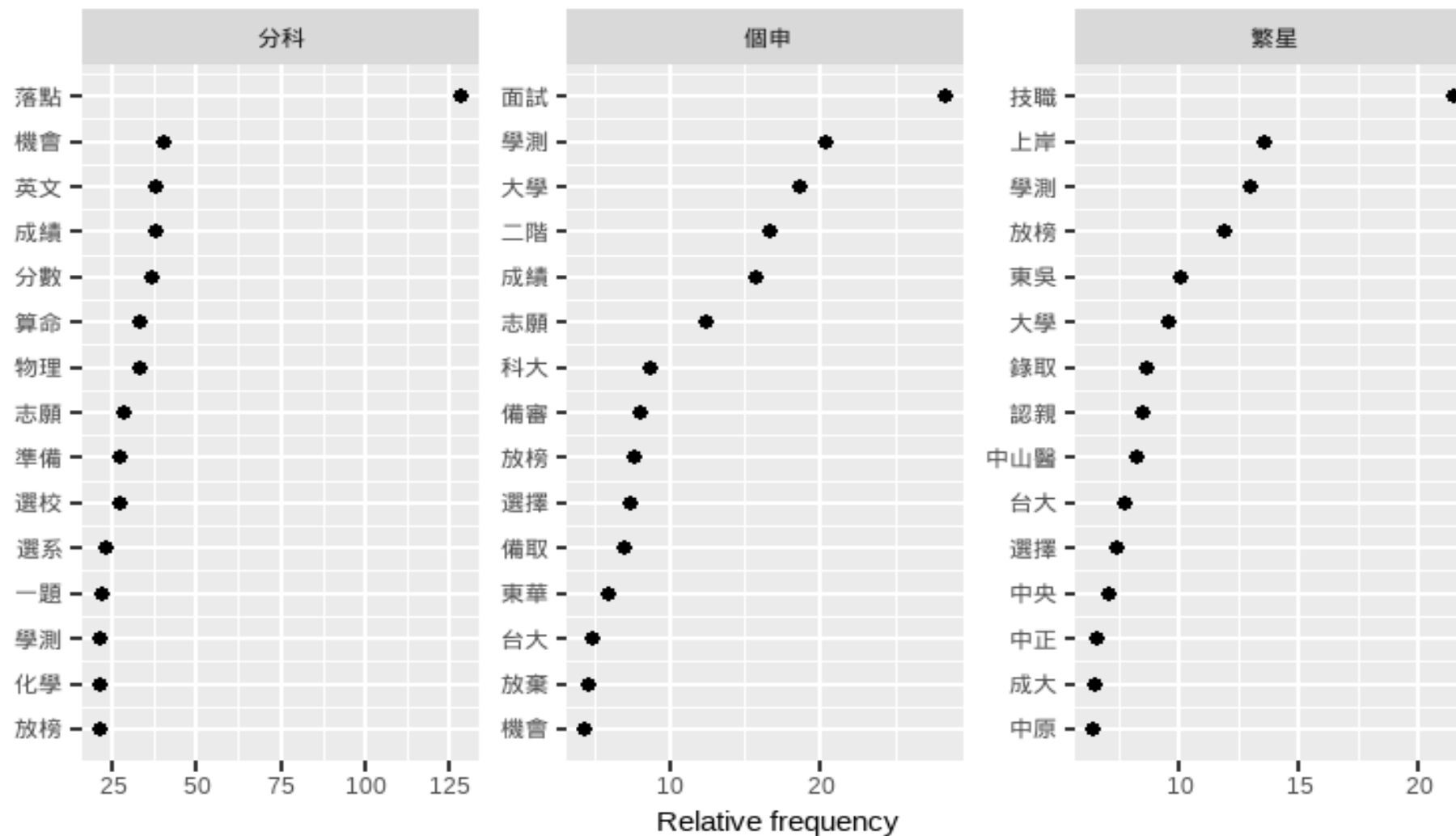
發文次數變化



跨年度之比較



各入學管道關鍵詞

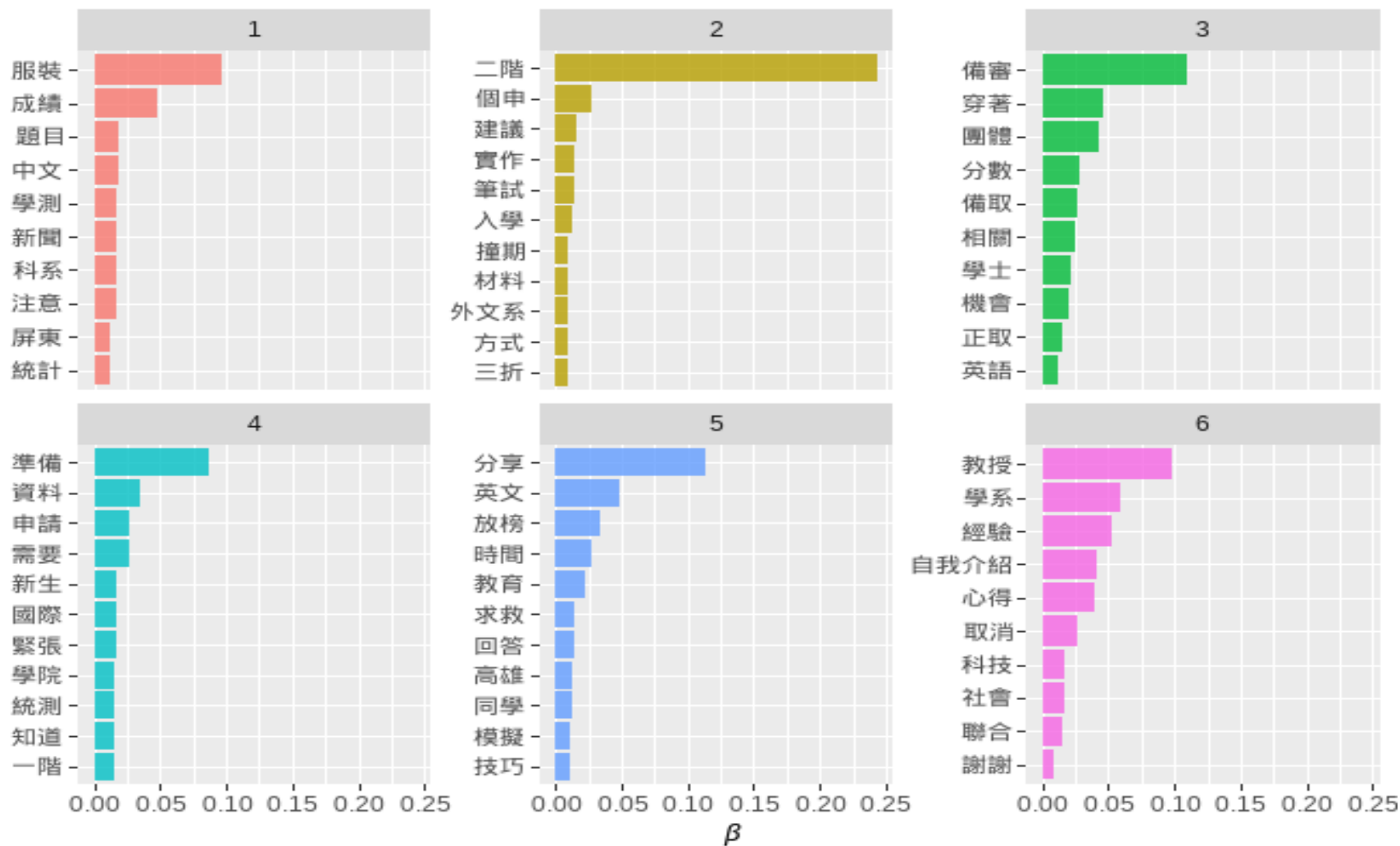


面試 - 主題 Topic Model

TOPIC 1	服裝、成績、題目
TOPIC 2	二階、個申、建議
TOPIC 3	備審、穿著、團體
TOPIC 4	準備、資料、申請
TOPIC 5	分享、英文、放榜
TOPIC 6	教授、學系、經驗

面試 - 各主題關鍵字

Highest word probabilities for each topic (interview)



結論與建議

- 結論：大學新生季板實際上是高三升學版
- 建議：停用詞與字典在該領域的適合程度

參考資料

- 一文詳解LDA主題模型
[url] : <https://zhuanlan.zhihu.com/p/31470216>
- 中文自然語言處理基礎
[url] : <https://ithelp.ithome.com.tw/articles/10192043>
- 文字探勘實驗室
[url] : <http://120.125.73.101/~asisweb/textcloud/nlp.php>
- 8 中文文本資料處理
[url] : <https://rlads2021.github.io/LabBook/ch08>
- [NLP][Python] 中文斷詞最方便的開源工具之一：Jieba
[url] : <https://clay-atlas.com/blog/2019/09/23/python-chinese-tutorial-tokenizer-jieba/>

Thanks for Listening ~ ~